

# Counting Process Based Dimension Reduction Methods for Censored Outcomes\*

Qiang Sun<sup>†</sup> Ruoqing Zhu<sup>‡</sup> Tao Wang<sup>§</sup> Donglin Zeng<sup>¶</sup>

April 18, 2017

## Abstract

We propose a class of dimension reduction methods for right censored survival data that exhibits significant numerical performances over existing approaches. The underlying framework of the proposed methods is based on a counting process representation of the failure process. Semiparametric estimating equations are constructed to estimate the dimension reduction subspace for the failure time model. Compared with existing approaches, the nonparametric component estimation in our framework is adaptive to the structural dimensions, and thus circumventing the curse of dimensionality. Asymptotic normality is established for the obtained estimators. Interestingly, with some intuitive heuristics,

---

\*The first two authors contribute equally.

<sup>†</sup>Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08540; e-mail: [qsun.ustc@gmail.com](mailto:qsun.ustc@gmail.com).

<sup>‡</sup>Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820; email: [rqzhu@illinois.edu](mailto:rqzhu@illinois.edu).

<sup>§</sup>Department of Bioinformatics and Biostatistics, Shanghai Jiao Tong University, Shanghai, China; email: [neowangtao@sjtu.edu.cn](mailto:neowangtao@sjtu.edu.cn).

<sup>¶</sup>Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; email: [dzeng@email.unc.edu](mailto:dzeng@email.unc.edu).

we further propose a computationally efficient singular value decomposition to directly estimate the dimension reduction subspace. No multidimensional non-parametric estimation is required. Extensive numerical studies and a real data analysis are conducted to demonstrate the performance of the proposed methodology.

**keywords** Counting Process, Dimension Reduction, Right Censoring, Semiparametric Inference, Survival Analysis.

## 1 Introduction

Dimension reduction has been a long-standing and important problem in regression analysis. It aims to extract a low-dimensional subspace from a  $p$ -dimensional space of covariates  $X = (X_1, \dots, X_p)^T$ , to predict an outcome of interest  $T$ . The dimension reduction literature often assumes the following multiple-index model

$$T = h(B^T X, \epsilon), \quad (1.1)$$

where  $\epsilon$  is a random error variable which is independent of  $X$ ,  $B \in \mathbb{R}^{p \times d}$  is a coefficient matrix with  $d < p$ , and  $h(\cdot)$  is a completely unknown link function. This model is equivalent to assuming  $T \perp X | B^T X$  (Li, 1991). Since any  $d$  linearly independent vectors in the linear space spanned by the columns of  $B$  also satisfy model (1.1) for some  $h$ , we define  $\mathcal{S}(B)$  to be this linear subspace. Furthermore, we call the intersection of all such subspaces satisfying  $T \perp X | B^T X$  as the central subspace, denoted by  $\mathcal{S}_{T|X}$ , and the dimension of  $\mathcal{S}_{T|X}$  is referred to as the structural dimension. According to Cook (2009),  $\mathcal{S}_{T|X}$  is uniquely defined under mild conditions. Hence, the goal of dimension reduction in (1.1) is to determine the structural dimension and the central subspace using empirical data.

There has been extensive literature on estimating the central subspace for completely observed data, including the seminal paper from Li (1991) and subsequent works such as

Cook and Weisberg (1991); Zhu et al. (2006); Li and Wang (2007); Xia (2007), and Ma and Zhu (2012), among many others. When  $T$  is subject to right censoring, which frequently occurred in survival analysis, model (1.1) includes many well-known survival models as special cases, for instance, the Cox proportional hazard model (Cox, 1972), the accelerated failure-time model (Lin et al., 1998), and the linear transformation models (Zeng and Lin, 2007). There has been a limited number of works on estimating the dimension reduction subspace using censored observations. Li et al. (1999) propose a modified sliced inverse regression method which uses the estimate of the conditional survival function to account for censored cases. Xia et al. (2010) propose to estimate the conditional hazard function, and then exploit the average directive estimation. Most of these existing methods need to estimate the conditional distribution of  $T$  or the censoring function given  $X$  nonparametrically, and will almost certainly suffer from the curse of dimensionality in data analysis (Xia et al., 2010). Additionally, some alternative approaches such as Lu and Li (2011) adopts an inverse probability weighting scheme, which implicitly requires the correct specification of the censoring mechanism.

In this paper, we propose a counting process based dimension reduction framework to estimate the dimension reduction subspace using right censored data. Our method differs from the existing methods in the following aspects. First, our approach is built upon a counting process representation of the conditional hazard. We utilize an interesting property that at each time point, the instantaneous failure process can be viewed as a binary outcome variable and the sliced inverse mean differences can be used. Second, it does not necessarily rely on the linearity assumption (Li, 1991), nor any other conditions on the covariates. This property makes our framework applicable in many general settings when  $X$  can be non-elliptical. Third, the proposed framework is adaptive to the structural dimension in the sense that the involved nonparametric estimation only depends on the dimension of  $\mathcal{S}(B)$ , and thus circumvents the curse of dimensionality. Furthermore, when we are willing

to make more restrict assumptions on  $X$ , our method reduces to a computationally efficient approach that can directly estimate the dimension reduction subspace without nonparametric smoothing. This resembles the simplicity of the original sliced inverse regression in [Li \(1991\)](#).

The paper proceeds as follows. Section 2 proposes a family of estimating equations for dimension reduction, based on a counting process representation. We further provide four specific examples that utilize these estimating equations to estimate  $B$ : a forward regression approach, two inverse regression approaches, and a computationally efficient one. The actual implementation and algorithms are given in Section 3. We provide the theoretical properties of the proposed method in Section 4. Extensive simulation studies and real data analysis are presented in Section 5 to demonstrate the practical utility of our proposed methods. Further discussions are presented in Section 6. Technical details are collected in the Appendix.

## 2 Proposed methods

### 2.1 Semiparametric estimating equations for the central subspace

Throughout the paper, we denote the failure time by  $T$  and denote the censoring time by  $C$ . Let  $Y = T \wedge C \equiv \min\{T, C\}$  and  $\delta = I(T \leq C)$  be the observed event time and the censoring indicator, respectively. We assume that  $C$  is independent of  $T$  conditional on  $X$ . Let  $N(u) = I(Y \leq u, \delta = 1)$  and  $Y(u) = I(T \wedge C \geq u)$  denote the observed counting process and at-risk process, respectively. Let  $\lambda(u|X)$  be the conditional hazard for  $T$  given  $X$ . Due to [Xia et al. \(2010\)](#), model (1.1) is equivalent to  $\lambda(u|X) = \lambda(u|B^T X)$ . We further let  $dM(u, X) = dM(u, B^T X) = dN(u) - \lambda(u|B^T X)Y(u)du$  be the martingale increment process indexed by  $u$ . The idea of this paper centers on constructing estimation equations that are built upon the counting process representation of the underlying survival model. To derive

the estimating equations, we follow the semiparametric analysis in [Bickel et al. \(1993\)](#) and [Tsiatis \(2007\)](#) (see derivations in [Appendix A](#)), and obtain the ortho-complement of the nuisance tangent space at  $B$ :

$$\mathcal{E}^\perp = \left\{ \int \{ \alpha(u, X) - \alpha^*(u, B^\top X) \} dM(u, X) : \forall \alpha(u, X) \right\}, \quad (2.1)$$

where  $\alpha(u, X)$  is an arbitrary functional curve of  $X$  that changes over time  $u$ , and

$$\alpha^*(u, B^\top X) = E\{ \alpha(u, X) | Y(u) = 1, B^\top X \}.$$

Then, to estimate the parameters  $B$ , we propose to use the following unbiased estimating functions

$$E \left[ \int \{ \alpha(u, X) - \alpha^*(u, B^\top X) \} \{ dN(u) - \lambda(u | B^\top X) Y(u) du \} \right] = 0, \quad (2.2)$$

while the sample version based on  $n$  i.i.d observations  $(Y_i, \Delta_i, X_i)$  is given by

$$n^{-1} \sum_{i=1}^n \left[ \int \{ \alpha(u, X_i) - \alpha^*(u, B^\top X_i) \} \{ dN_i(u) - \lambda(u | B^\top X_i) Y_i(u) du \} \right] = 0, \quad (2.3)$$

where the conditional hazard function needs to be estimated using the data.

It is then crucial to choose specific forms of  $\alpha(u, X)$  properly. Different choices may result in simplifications of the formulation and/or may gain additional theoretical and computational advantages. In the next two sections, we present four different choices, which fall into two major categories: the forward and inverse regression schemes. The main differences between the two schemes is whether the counting process  $N(u)$  is used in the definition of  $\alpha(u, X)$ . The forward regression scheme is essentially the estimating equations approach in the normal regression, while the inverse regression scheme utilize  $N(u)$  to mimic the sliced inverse regression ([Li, 1991](#)) conceptually.

## 2.2 Forward regression scheme

In the forward regression scheme, we choose  $\alpha(u, X)$  such that it does not depend on the observed failure process  $N(u)$ . We first notice that, as long as  $\alpha(u, X)$  depends at most on

the at-risk process  $Y(u)$ , the following property holds (see Appendix A.1 for details):

$$\alpha^*(u, B^T X) = \frac{E\{\alpha(u, X)Y(u)|B^T X\}}{E\{Y(u)|B^T X\}} = E\{\alpha(u, X)|Y(u) = 1, B^T X\},$$

We can then further simplify the estimating functions by noticing that

$$\begin{aligned} 0 &= E\left(\int \left[\alpha(u, X) - E\{\alpha(u, X)|Y(u) = 1, B^T X\}\right] dM(u, X)\right) \\ &= E\left(\int \left[\alpha(u, X) - E\{\alpha(u, X)|Y(u) = 1, B^T X\}\right] dN(u)\right). \end{aligned} \quad (2.4)$$

This formulation resembles the estimating functions in the Cox proportional hazard model. The forward regression scheme is not the focus of this paper, and we shall give only one example of  $\alpha(u, X)$  for demonstration. It concerns the situation when one has a strong prior belief that the underlying structural dimension is  $d = 1$ . We also note that this approach require only a 1-dimensional nonparametric estimation.

**Example 2.1** (FR). Let  $\alpha(u, X) = X$  and the population version of the  $p$ -dimensional estimating equations is given by:

$$E\left(\int \left[X - E\{X|Y(u) = 1, B^T X\}\right] dN(u)\right) = 0. \quad (2.5)$$

This formulation reduces to the efficient estimating equations for Cox proportional hazard model when the exponential link is known to be the underlying truth. It can be also used for transformation models instead of the NPMLE proposed by Zeng and Lin (2007). Based on our numerical experience, the method has the best performance if the true structural dimension  $d$  is 1. However, it will not be able to handle cases with  $d > 1$ .

**Remark 2.2.** We note that some simple extension of this method, such as letting  $\alpha(u, X) = E\{XY(u)|X\}$ , shall lead to  $p$ -by- $p$  dimensional estimating equations and will be capable to deal with higher structural dimension  $d > 1$ . Then a  $d$ -dimensional nonparametric estimation is then needed in such extensions.

To implement the method above, we utilize the generalized method of moments (GMM) approach ([Hansen, 1982](#)):

$$\beta = \arg \min_{\beta \in \Theta} \left\{ \psi_n(\beta)^\top \psi_n(\beta) \right\}, \quad (2.6)$$

where  $\psi_n(\beta)$  are sample means of the moment conditions, and  $\beta = \text{vec}(B)$  is the vector concatenating of the columns of  $B$ . Noticing that  $dN_i(u)$  takes a jump at the  $Y_i$  only if  $\delta_i = 1$ , we can estimate  $\psi_n(\beta)$  in (2.5) by

$$\hat{\psi}_n(\beta) = \frac{1}{n} \sum_{i=1}^n \{X_i - \hat{E}(X|Y > Y_i, B^\top X_i)\} \delta_i, \quad (2.7)$$

where  $\hat{E}(X|Y > u, B^\top X = z)$  takes the following forms for any given  $u$  and  $z$ ,

$$\frac{\sum_{i=1}^n X_i I(Y_i > u) K_h(B^\top X_i - z)}{\sum_{i=1}^n I(Y_i > u) K_h(B^\top X_i - z)}.$$

In the above nonparametric estimations of the conditional expectations,  $K_h(\cdot) = h^{-1}K(\cdot/h)$  for any bandwidth  $h$  and  $K(\cdot)$  is a kernel function. The choice of the bandwidth  $h$ , the kernel  $K(\cdot)$  and the numerical approach for solving the estimating equations are deferred to [Section 3](#).

### 2.3 Inverse regression scheme

In this section, we focus the main proposal of this paper: the inverse regression scheme. We will use  $N(u)$  in the construction of  $\alpha(u, X)$ , which conceptually resembles the sliced inverse regression ([Li, 1991](#)). Note that all the proposed method in this section will construct  $p$ -by- $p$  dimensional estimating equations and be able to handle structural dimension  $d > 1$ . An important property that motivates our method is

$$\{dN(u) | Y(u) = 1, B^\top X\} \sim \text{Bernoulli}\{\lambda(u|B^\top X)du\}, \quad (2.8)$$

where  $dN(t) = N(t + dt) - N(t)$ . This suggests that at any time point  $u$ , the failure indicator of an at-risk subject follows the bernoulli distribution. Hence, we can utilize the

sliced conditional mean of  $X$  given the outcome of  $dN(t)|Y(t) = 1$ . This leads to the construction of a local mean difference that is essentially the sliced mean difference for the binary outcome  $dN(u)$  (Cook and Lee, 1999):

$$\varphi(u) = E(X|dN(u)=1, Y(u)=1) - E(X|dN(u)=0, Y(u)=1) \quad (2.9)$$

It should be noted that  $\varphi(u)$  is a constant given any time point  $u$ , and it is within the central subspace  $\mathcal{S}_{T|X}$  since the outcome  $dN(u)$  conditioning on the event  $Y(u) = 1$  depends only on the failure model  $\lambda(u|B^T X)$  (Xia et al., 2010). We consider three forms of the estimating equations, which all uses the following construction of  $\alpha(u, X)$ :

$$\alpha(u, X) = X\varphi^T(u). \quad (2.10)$$

Noticing that  $\varphi^T(u)$  is a constant term at any  $u$ , we have

$$\alpha(u, X) - \alpha^*(u, B^T X) = \{X - E(X|Y(u) = 1, B^T X)\}\varphi^T(u). \quad (2.11)$$

This, together with estimating functions constructed in (2.2), leads to three proposed methods, presented in the following.

**Example 2.3** (IR-Semi). Replacing  $\alpha(u, X) - \alpha^*(u, B^T X)$  in (2.2) by (2.11) leads to the inverse regression semiparametric approach (IR-Semi) in its population version:

$$E\left[\int \{X - E(X|Y(u) = 1, B^T X)\}\varphi^T(u)dM(u)\right]. \quad (2.12)$$

Note again that (2.12) are  $p$ -by- $p$  dimensional estimating functions, and is able to handle  $d > 1$ , however, the nonparametric estimation part is only  $d$  dimensional as reflected by  $B^T X$ . Furthermore, this formulation enjoys the double robustness property which is illustrated in Appendix A.2. Similar phenomenon has been observed by Ma and Zhu (2012) in the regression setting but without censoring. This suggests that if the term  $E(X|Y(u) = 1, B^T X)$  or  $M(u)$  are estimated incorrectly, we can still obtain consistent estimations of the



dimension reduction subspace. In our numerical experiment, we do observe some numerical advantage of IR-Semi over its simplified version IR-CP, which is given in the next example.

To implement the IR-Semi approach, we again exploit the GMM approach used in (2.6), with

$$\hat{\psi}_n(\beta) = \text{vec} \left[ \frac{1}{n} \sum_{i=1}^n \sum_{\substack{j=1 \\ \delta_j=1}}^n \left\{ X_i - \hat{E}(X|Y \geq Y_j, B^T X_i) \right\} \hat{\varphi}^T(Y_j) \left\{ \delta_i I(j=i) - \hat{\lambda}(Y_j|B^T X_i) \right\} \right] \quad (2.13)$$

where

$$\hat{E}(X|Y \geq u, B^T X = z) = \frac{\sum_{i=1}^n X_i K_h(B^T X_i - z) I(Y_i \geq u)}{\sum_{i=1}^n K_h(B^T X_i - z) I(Y_i \geq u)}, \quad (2.14)$$

$\varphi(u)$  can be estimated by the sliced average based on (2.9), for some bandwidth  $h$ :

$$\hat{\varphi}(u) = \frac{\sum_{i=1}^n X_i I(u \leq Y_i < u + h, \delta_i = 1)}{\sum_{i=1}^n I(u \leq Y_i < u + h, \delta_i = 1)} - \frac{\sum_{i=1}^n X_i I(Y_i \geq u)}{\sum_{i=1}^n I(Y_i \geq u)}. \quad (2.15)$$

Furthermore, following Dabrowska et al. (1989), the conditional hazard function at any time point  $u$  is estimated by

$$\hat{\lambda}(u|B^T X = z) = \frac{\sum_{i=1}^n I(Y_i = u) I(\delta_i = 1) K_h(B^T X_i - z)}{\sum_{i=1}^n I(Y_i \geq u) K_h(B^T X_i - z)}.$$

**Example 2.4** (IR-CP). Similar to the forward regression example, our choice of  $\alpha(u, X)$  in Equation (2.10) depends on at most the at-risk process  $Y(u)$ . Hence, the estimating functions defined in CP-Semi (2.12) can be simplified to the following counting process version of the inverse regression (IR-CP):

$$E \left[ \int \{X - E(X|Y(u) = 1, B^T X)\} \varphi^T(u) dN(u) \right]. \quad (2.16)$$

Replacing  $dM(u)$  with  $dN(u)$  dramatically reduces the computational burden. This can be seen from Equation (2.13), where a conditional hazard function  $\hat{\lambda}(Y_j|B^T X_i)$  needs to be evaluated at each observed failure time point  $j$  for all observations  $i$ . In contrast, the

formulation in (2.16) does not require the computation of  $\hat{\lambda}(Y_i|B^T X_i)$ . Of course, by doing this, we lose the double robustness property.

The implementation of the IR-CP approach is a simplified version of Equation (2.13):

$$\hat{\psi}_n(\beta) = \text{vec} \left[ \frac{1}{n} \sum_{i=1}^n \left\{ X_i - \hat{E}(X|Y \geq Y_i, B^T X_i) \right\} \hat{\varphi}^T(Y_i) \right] \quad (2.17)$$

where the quantities  $\hat{E}(X|Y > u, B^T X_i = z)$  and  $\hat{\varphi}^T(u)$  are already given in Equations (2.14) and (2.15) respectively.

**Example 2.5** (CPSIR). In our last example, we explore a computationally efficient approach. Although the formulation in the IR-CP approach (Equation (2.17)) requires only a  $d$ -dimensional kernel estimation for the quantity  $\hat{E}(X|Y \geq Y_i, B^T X_i)$ , the solution can only be obtained through numerical optimization. To the best of our knowledge, there is no approach in the dimensional reduction literature for survival analysis that does not require any multidimensional kernel estimation or any plug-in estimator (e.g. the inverse probability of censoring weights used in Lu and Li (2011)). Hence, a computationally efficient approach that only requires a singular value decomposition (SVD) has a great advantage. Of course, this comes at a price of making additional assumptions, which needs to be further investigated, but our numerical results suggest that it outperforms all existing methods, even when the assumptions are violated.

Noticing that in the formulation (2.16), the major computational cost is to estimate  $E(X|Y(u) = 1, B^T X)$  for each observation, if we make the following *local linearity* and *local constant variance condition*, the problem can be much simplified.

**Definition 2.6** (Local Linearity Condition and Local Proportional Constant Variance Condition). For any  $\alpha \in \mathbb{R}^p$  and any  $u > 0$ , the linearity condition (Li, 1991) is satisfied further conditioning on the event  $\{Y(u) = 1\}$ , i.e.,

$$E(\alpha^T X|Y(u) = 1, B^T X = z) = c_0(u) + c^T(u)z \quad (2.18)$$

where  $c_0(u)$  and  $c(u)$  are constants that possibly depend on  $u$ . Furthermore, the local proportional constant variance condition is a parallel version of [Cook and Weisberg \(1991\)](#), that requires

$$\text{cov}(X|Y(u) = 1, B^T X) = c_1(u)\Sigma \quad (2.19)$$

where  $c_1(u)$  is some constant that depends on  $u$ .

Based on these two assumption, we can further simplify Equation (2.16). Noticing that after centering  $X$  at time point  $u$ ,

$$\begin{aligned} & E(X|Y(u) = 1, B^T X) - E(X|Y(u) = 1) \\ &= P[X - E(X|Y(u) = 1)], \end{aligned}$$

where  $P = \Sigma B(B^T \Sigma B)^{-1} B^T$ . Then the estimating equations reduce to

$$Q E \left[ \int \{X - E(X|Y(u) = 1)\} \varphi^T(u) dN(u) \right] = 0, \text{ where } Q = I - P,$$

which are equivalent to deriving the left-singular space of the covariance matrix

$$E \left[ \int \{X - E(X|Y(u) = 1)\} \varphi^T(u) dN(u) \right]. \quad (2.20)$$

The computation of this approach is extremely simple. Realizing that  $dN(u)$  takes value 1 at at most one time point on the entire time domain, which corresponds to the failure subjects, the covariance form can be estimated by a sum of  $n$  terms, where no kernel estimation is required. Then we perform SVD on this sample covariance matrix and obtain its leading left singular vectors. Details are provided in [Algorithm \(1\)](#).

**Remark 2.7.** We note that the two local conditions are somehow restrictive, and does not always hold. For example, since  $Y(u)$  is a process that depends on both the failure and censoring distribution, as long as the censoring distribution depends on structures beyond  $B^T X$ , this condition could be violated. However, many recent literatures argue

that the sliced inverse regression seems to still have satisfactory performances even when the linearity condition is violated (Li and Dong, 2009; Dong and Li, 2010). Hence, this does not prevent the CPSIR method from serving as a good explorative tool even these conditions are not satisfied. Practically, it can also be used as the initial value when solving our other approaches to speed up the computation.

### 3 Implementation and Algorithm

The implementation of the CPSIR method is in fact straightforward since only sliced averaging and eigen-decomposition are required. The numerical approach is highlighted in Example 2.5. We provide the details in the following Algorithm (1).

---

**Algorithm 1** The SVD algorithm for CPSIR.

---

- 1: **Algorithm:**  $\widehat{B} \leftarrow \text{CPSIR}(\{(X_i, \delta_i, Y_i), 1 \leq i \leq n\}, h, k)$
  - 2: **Input:**  $\{(X_i, \delta_i, Y_i), 1 \leq i \leq n\}, h > 0, k > 0$
  - 3: **Do**
  - 4:   For each  $Y_i$  such that  $\delta_i = 1$ , i.e., each observed failure time, calculate  $\widehat{\varphi}(Y_i)$  using Equation (2.15) and  $\widehat{E}(X|Y > Y_i)$  using
$$\widehat{E}(X|Y > u) = \left\{ \sum_{i=1}^n I(Y_i > u) \right\}^{-1} \left\{ \sum_{i=1}^n X_i I(Y_i > u) \right\}.$$
  - 5:   Calculate  $\widehat{M} = n^{-1} \sum_{\delta_i=1} \{X_i - \widehat{E}(X|Y_i)\} \widehat{\varphi}^T(Y_i)$
  - 6:   Perform SVD:  $\widehat{M} = \widehat{U} \widehat{D} \widehat{V}^T$
  - 7: **Return**  $\widehat{B}$  as the first  $k$  columns of  $\widehat{U}$ .
- 

The forward regression and the two inverse regression approaches IR-Semi and IR-CP all require numerical optimization to solve the estimating equations. Using the corresponding choices of the moment conditions specified in (2.7), (2.13) and (2.17), we solve for the minimizer of  $\widehat{\psi}_n(\beta)^T \widehat{\psi}_n(\beta)$ , where  $\beta = \text{vec}(B)$  is the vector concatenating of the columns of  $B$ . Hence, general-purpose optimization tools such as the Newton–Raphson method can

be used. At each iteration, we compute the gradient vector  $\nabla_{\beta} \|\psi_n(\beta)\|_2^2$  and the Hessian matrix  $\nabla_{\beta\beta} \|\psi_n(\beta)\|_2^2$  evaluated at the current  $\beta$  value. The working parameter vector is then updated by

$$\beta^{(k+1)} = \beta^{(k)} - \left\{ \nabla_{\beta\beta} \|\psi_n(\beta^{(k)})\|_2^2 \right\}^{-1} \nabla_{\beta} \|\psi_n(\beta^{(k)})\|_2^2.$$

for each interaction  $k$ . The full algorithm is presented in Algorithm 2. The iteration is stopped when a pre-specified optimization precision  $\varepsilon_0$  is reached.

---

**Algorithm 2** The Newton-Raphson algorithm for solving the GMM approach for inverse regression.

---

- 1: **Algorithm:**  $\hat{\beta} \leftarrow \text{Newton-Raphson}(\varepsilon_0, \beta^{(0)}, \{(X_i, \delta_i, Y_i), 1 \leq i \leq n\})$
  - 2: **Input:**  $\varepsilon_0, \beta^{(0)}, \{(X_i, \delta_i, Y_i), 1 \leq i \leq n\}$
  - 3: **Initialize:**  $\beta^{(0)} \leftarrow 0, k = 0$
  - 4: **Repeat**
  - 5:   Calculate  $\nabla_{\beta} \|\psi_n(\beta^{(k)})\|_2^2$  and  $\nabla_{\beta\beta} \|\psi_n(\beta^{(k)})\|_2^2$  approximately
  - 6:    $\beta^{(k+1)} \leftarrow \beta^{(k)} - \left\{ \nabla_{\beta\beta} \|\psi_n(\beta^{(k)})\|_2^2 \right\}^{-1} \nabla_{\beta} \|\psi_n(\beta^{(k)})\|_2^2, \quad k \leftarrow k + 1$
  - 7: **Until**  $\|\beta^{(k+1)} - \beta^{(k)}\|_2 \leq \varepsilon_0$
  - 8: **Return**  $\hat{\beta} = \beta^{(k+1)}$
- 

In our numerical implementations, we exploit the Gaussian kernel and choose the optimal bandwidth  $h = (4/(d+2))^{1/(d+4)} n^{-1/(d+4)} \hat{\sigma}$  (Silverman, 1986), where  $\hat{\sigma}$  is the estimated standard deviation. To prevent the possible non-identifiability issue of  $B$  (Ma and Zhu, 2012), we add a penalty,  $100 \cdot \|B^T B - I_{d \times d}\|_F$ , where  $\|\cdot\|_F$  denotes the Frobenius norm, on the right hand side of (2.6) to force orthogonality between columns of  $B$ .

## 4 Asymptotical Normality

We assume that the estimator  $\widehat{B}$  is achieved by solving the following semiparametric estimating equations

$$\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \widehat{\alpha}^*(u, \widehat{B}^\top X_i) \right\} d\widehat{M}(u, \widehat{B}^\top X_i) \right] = 0,$$

which covers both the inverse and forward regression schemes. We need the the following regularity assumptions.

**Assumption 4.1.** There exists a  $\tau$ , such that  $0 < \tau < \infty$  and  $\text{pr}(Y > \tau | X) > 0$ .

**Assumption 4.2.** For any  $i = 1, \dots, n$ , the conditional cumulative hazard evaluated at  $\tau$ ,  $\Lambda_i(\tau)$ , is bounded.

**Assumption 4.3.** There exists a predictable and square-integrable function  $\Gamma(u)$  with respect to  $\Lambda_i(u)$ ,  $i = 1, \dots, n$ , such that, for any  $X$  and  $\widehat{B} \in \Omega(B)$  and  $0 \leq u \leq \tau$ , we have

$$\begin{aligned} \left| \text{vec} \{ \alpha^*(u, \widehat{B}^\top X) \} - \text{vec} \{ \alpha^*(u, B^\top X) \} - [\nabla_\beta \text{vec} \{ \alpha^*(u, B^\top X) \}] (\widehat{\beta} - \beta) \right| &\leq \Gamma(u) \|\widehat{\beta} - \beta\|_2^2, \\ \text{and } \left| \lambda(u, \widehat{B}^\top X) - \lambda(u, B^\top X) - \{ \nabla_\beta \lambda(u, B^\top X) \}^\top (\widehat{\beta} - \beta) \right| &\leq \Gamma(u) \|\widehat{\beta} - \beta\|_2^2. \end{aligned}$$

**Assumption 4.4.** We assume that for some  $\kappa < 1/2$ , the following uniform rate of convergence holds for the conditional nonparametric estimation:

$$\begin{aligned} \sup_{0 \leq u \leq \tau, \widehat{B} \in \Omega(B)} \left\| \text{vec} \left\{ \widehat{\alpha}^*(u, \widehat{B}^\top X) - \alpha^*(u, \widehat{B}^\top X) \right\} \right\|_2 &= O_p(n^{-1/2+\kappa/2}), \quad \text{and} \\ \sup_{0 \leq u \leq \tau, \widehat{B} \in \Omega(B)} \left\| \widehat{\lambda}(u, \widehat{B}^\top X) - \lambda(u, \widehat{B}^\top X) \right\|_2 &= O_p(n^{-1/2+\kappa/2}). \end{aligned}$$

**Assumption 4.5.** The conditional cumulative hazard satisfies  $\widehat{\Lambda}(t, B^\top X) - \Lambda(t, B^\top X) = O_p(n^{-1/2})$ .

Assumptions 4.1-4.2 are standard in survival analysis. Assumption 4.3 basically requires that  $\alpha^*(u, \cdot)$  and  $\lambda(u, \cdot)$  are smooth enough. Assumption 4.4 concerns the rate of convergence

for the estimates of  $\alpha^*$  and  $\lambda$ . The rate can be slower than root- $n$  and is achievable by most of the nonparametric regression methods. Assumption 4.5 requires that the cumulative hazard have the root- $n$  rate of convergence. This can be achieved by using the integrated kernel estimator, see Delyon et al. (2016). We now present our main theorem, with proof deferred to the appendix.

**Theorem 4.6** (Asymptotic Normality). Under Assumptions 4.1-4.5, the estimator  $\hat{B}$  is asymptotically normal, that is  $\text{vec}(\hat{B} - B) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ , for some positive definite covariance matrix  $\Sigma$ .

## 5 Numerical Examples

### 5.1 Simulation Studies

In this section, we examine the finite sample performance of our proposed estimating schemes via extensive numerical experiments. Specifically, we carry out the estimation of dimension reduction subspace using both the forward and inverse regression. For the forward regression approach, FR is defined by estimating equations (2.5) in Example 2.1. For the inverse regression approaches, the fully semiparametric approach is carried out using estimating equations (2.12) in Example 2.3, and the computational efficient approach is carried out using estimating equations (2.20) in Example 2.5.

We compare our methods with three existing approaches: the double slicing (DS) approach in Li et al. (1999); the minimal average variance estimation based on hazard functions (hMave) in Xia et al. (2010); and the inverse probability of censoring weighted approach (SIR-ipcw) proposed by Lu and Li (2011). For DS, we use the R package “censorSIR” provided by Sun and Wu (2006). The hMave implementation was provided by the original author, and we carry out the SIR-ipcw by ourselves while utilizing the “dr” package (Weisberg, 2002).

Four different settings are considered: Setting 1 is a classical Cox proportional hazard model; Setting 2 is constructed with structural dimension  $d=2$  and directions in the hazard function are changing over time; Setting 3 also has structural dimension equal to two, with the two directions interacting with each other. Setting 4 also has two interacting structural dimensions, while the failure and the censoring variables have overlap. For each setting we consider the total number of dimension  $p$  being 6 and 12. Each experiment is repeated 200 times with sample size  $n=400$ .

Setting 1: The true survival time  $T$  and the censoring time  $C$  are generated from exponential distributions with rate  $\exp(\beta^T X)$  and  $\exp(X_4 + X_5 - 1)$  respectively, where  $\beta = (1, 0.5, 0, \dots, 0)^T$  and  $X_j$  is the  $j$ -th element of  $X$ , for  $1 \leq j \leq p$ . The covariate  $X$  follows from multivariate normal distribution with mean 0 and covariance  $\Sigma = (0.5^{|i-j|})_{ij}$ . The overall censoring rate is around 35.3%.

Setting 2: We generate  $T_1$  and  $T_2$  from exponential distributions with rate  $\exp(\beta_1^T X)$  and  $\exp(\beta_2^T X)$  respectively, where  $\beta_1 = (1, 0, 1, 0, \dots, 0)^T$  and  $\beta_2 = (0, 1, 0, 1, 0, \dots, 0)^T$ . The true survival time  $T = T_1 I(T_1 < 0.4) + (T_2 + 0.4) I(T_1 \geq 0.4)$ . The censoring time  $C$  is generated from exponential distributions with rate  $\exp(X_5 - X_6 - 2)$ . The covariate  $X$  follows the same distribution as in Setting 1. The overall censoring rate is around 35.1%.

Setting 3: The true survival time  $T$  is generated from Weibull distribution with shape parameter 5 and scale parameter  $\exp(4\beta_2^T X(\beta_1^T X - 1))$ , where  $\beta_1 = (1, 0, 1, 0, \dots, 0)^T$  and  $\beta_2 = (0, 1, 0, 1, 0, \dots, 0)^T$ . The censoring time  $C$  is generated uniformly from 0 to  $3 \exp(X_5 - X_6 + 0.5)$ . We further draw  $X$  such that  $X_j$ 's follow standard uniform distribution  $U(0, 1)$  independently. The overall censoring rate is around 33.8%.

Setting 4: The true survival time  $T$  is generated from a Cox proportional hazard model with  $\log(T) = -0.25 + \beta_1^T X + 0.5\beta_1^T X\beta_2^T X + 0.25 \log(-\log(1 - u))$  and  $\log(C) = -0.5 + \beta_3^T X + \log(-\log(1 - u))$ , where  $u$ 's are i.i.d. uniform distributed,  $\beta_1 = (1, 1, 0, \dots, 0)^T$ ,  $\beta_2 = (0, 0, 1, -1, 0, \dots, 0)^T$ , and  $\beta_3 = (0, 1, 0, 1, 1, 1, 0, \dots, 0)^T$ . The covariate  $X$  follows the



Table 1: Setting 1 with structural dimension  $d = 1$ . We report the mean and standard deviations (in parenthesis) under three different measures.

Method	Frobenius	Trace	Canonical	Frobenius	Trace	Canonical
	predictor dimension $p = 6$			predictor dimension $p = 12$		
DS	0.31 (0.10)	0.95 (0.03)	0.99 (0.01)	0.44 (0.10)	0.90 (0.05)	0.97 (0.01)
hMave	0.67 (0.13)	0.77 (0.09)	0.86 (0.06)	0.73 (0.11)	0.73 (0.08)	0.85 (0.05)
SIR-ipcw	0.49 (0.16)	0.87 (0.09)	0.96 (0.03)	0.65 (0.16)	0.78 (0.13)	0.93 (0.09)
FR	0.22 (0.07)	0.97 (0.02)	0.99 (0.00)	0.32 (0.07)	0.95 (0.02)	0.99 (0.01)
CP-SIR	0.27 (0.09)	0.96 (0.03)	0.99 (0.01)	0.39 (0.09)	0.92 (0.04)	0.98 (0.01)
IR-Semi	0.24 (0.07)	0.97 (0.02)	0.99 (0.00)	0.35 (0.08)	0.93 (0.03)	0.99 (0.01)
IR-CP	0.23 (0.07)	0.97 (0.02)	0.99 (0.00)	0.33 (0.08)	0.94 (0.03)	0.99 (0.01)

same distribution as setting 1, except  $\Sigma = (0.25^{|i-j|})$ . The overall censoring rate is around 26.2%.

We investigate the statistical performance using three different measures: the Frobenius norm distance between the projection matrix  $P$  and its estimator  $\hat{P}$ , in which,  $P = B(B^T B)^{-1} B^T$ ; the trace correlation  $\text{tr}(P\hat{P})/d$ , where  $d$  is the structural dimension; and the canonical correlation between  $B^T X$  and  $\hat{B}^T X$ . The results are summarized in Tables 1-4.

Overall, the two inverse regression methods (IR-CP and IR-Semi) achieve the best performance, followed by the computationally efficient approach (CP-SIR). It is worth to point out that the CP-SIR method, while no non-parametric approximation is required, outperforms existing methods in all settings. Among all competing methods, DS performs the best, while hMave and SIR-ipcw outperforms DS occasionally. In terms of the three error measurements, we found that the Frobenius is the most informative measurement, while the Trace and Canonical correlations are less sensitive to the performances.

IR-CP and IR-Semi perform similarly while IR-Semi is slightly better in Settings 2 and 4. The main advantage of the IR-Semi approach compared with IR-CP is the double

Table 2: Setting 2 with structural dimension  $d = 2$ . We report the mean and standard deviations (in parenthesis) under three different measures.

Method	Frobenius	Trace	Canonical	Frobenius	Trace	Canonical
	predictor dimension $p = 6$			predictor dimension $p = 12$		
DS	0.46 (0.13)	0.94 (0.03)	0.97 (0.02)	0.71 (0.14)	0.87 (0.05)	0.94 (0.03)
hMave	1.18 (0.30)	0.63 (0.16)	0.72 (0.16)	1.40 (0.18)	0.50 (0.12)	0.64 (0.12)
SIR-ipcw	0.58 (0.20)	0.91 (0.07)	0.96 (0.05)	0.85 (0.20)	0.81 (0.09)	0.91 (0.06)
FR	0.67 (0.21)	0.88 (0.07)	0.91 (0.06)	0.89 (0.14)	0.80 (0.06)	0.86 (0.05)
CP-SIR	0.38 (0.11)	0.96 (0.03)	0.98 (0.01)	0.62 (0.12)	0.90 (0.04)	0.95 (0.02)
IR-Semi	0.35 (0.10)	0.97 (0.02)	0.98 (0.01)	0.65 (0.15)	0.89 (0.05)	0.94 (0.03)
IR-CP	0.36 (0.12)	0.96 (0.02)	0.98 (0.02)	0.67 (0.19)	0.88 (0.07)	0.94 (0.04)

robustness, which ensures consistency even when the conditional expectations are not estimated correctly. However, this theoretical advantage does not seem to translate into strong numerical improvements. This is possibly due to the variations in the hazard function estimation, which introduces less numerical stability. The main disadvantage of the IR-Semi approach is the computational cost. Under our settings with  $d = 2$ , it usually takes over an hour for the optimization to converge (using R package “optimx”), while the IR-CP approach requires only a few minutes. Based on our experiments, we advocate the use of IR-CP in practice, while IR-Semi can be an alternative when there is no constraint on the computational power.

In setting 1, FR achieves the best overall performance. As discussed in Example 1, FR mimics the efficient estimating equations used in the Cox proportional hazard model and is thus the most efficient method in this setting. In setting 2, The CP-SIR approach performs similarly to the two inverse regression approaches, and it even outperforms them under  $p = 12$ . This shows some potential of this approach in higher dimension settings when nonparametric estimations may not be preferred.

Table 3: Setting 3 with structural dimension  $d = 2$ . We report the mean and standard deviations (in parenthesis) of three estimation errors.

Method	Frobenius	Trace	Canonical	Frobenius	Trace	Canonical
	predictor dimension $p = 6$			predictor dimension $p = 12$		
DS	0.41 (0.14)	0.95 (0.03)	0.99 (0.01)	0.63 (0.13)	0.90 (0.04)	0.97 (0.02)
hMave	0.39 (0.16)	0.96 (0.04)	0.99 (0.03)	0.67 (0.25)	0.87 (0.10)	0.95 (0.07)
SIR-ipcw	0.68 (0.19)	0.87 (0.07)	0.97 (0.02)	0.94 (0.17)	0.77 (0.08)	0.94 (0.05)
FR	1.07 (0.13)	0.71 (0.07)	0.87 (0.07)	1.10 (0.11)	0.70 (0.06)	0.84 (0.11)
CP-SIR	0.36 (0.13)	0.96 (0.03)	0.99 (0.01)	0.55 (0.11)	0.92 (0.03)	0.98 (0.02)
IR-CP	0.22 (0.09)	0.99 (0.01)	1.00 (0.00)	0.40 (0.14)	0.96 (0.04)	0.99 (0.01)
IR-Semi	0.27 (0.12)	0.98 (0.02)	1.00 (0.01)	0.51 (0.16)	0.93 (0.05)	0.98 (0.02)

Setting 3 is an example where hMave outperforms DS. This is possible due to the uniform distribution of the features which violates the linearity assumptions in DS. As contrast, even the assumptions in CP-SIR is severely violated, the performance is still satisfactory. Again, the two inverse approaches achieve significantly better performances. Setting 4 is a similar case to setting 3, while the covariates are generated from normal distributions. SIR-ipcw slightly outperforms DR in this case possibly due to the consistent estimation of the censoring weights. CP-SIR slightly outperforms SIR-ipcw, however, the advantage of IR-CP and IR-Semi is significant.

## 5.2 Skin Cutaneous Melanoma Data Analysis

We apply the proposed method to The Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov/>) skin cutaneous melanoma (SKCM) dataset. TCGA provides the public with one of the most comprehensive profiling data on more than thirty cancer types. We acquire gene expression and clinical data on a total of 469 patients (156 observed failures) and their mRNA expression data on 20,531 genes. To produce biologically meaningful results, we preselect 20 genes in this analysis, which are the top 20 genes highly associated with cu-

Table 4: Setting 4 with structural dimension  $d = 2$ . The mean and standard deviations (in parenthesis) of three estimation errors are reported.

Method	Frobenius	Trace	Canonical	Frobenius	Trace	Canonical
	predictor dimension $p = 6$			predictor dimension $p = 12$		
DS	0.46 (0.11)	0.94 (0.03)	0.96 (0.02)	0.61 (0.12)	0.90 (0.04)	0.94 (0.03)
hMave	1.42 (0.04)	0.50 (0.03)	0.59 (0.04)	1.46 (0.07)	0.47 (0.06)	0.56 (0.07)
SIR-ipcw	0.39 (0.16)	0.96 (0.05)	0.97 (0.05)	0.59 (0.17)	0.91 (0.06)	0.95 (0.05)
FR	1.23 (0.10)	0.62 (0.06)	0.72 (0.06)	1.18 (0.11)	0.65 (0.06)	0.76 (0.05)
CP-SIR	0.34 (0.08)	0.97 (0.01)	0.98 (0.01)	0.50 (0.09)	0.93 (0.02)	0.97 (0.01)
IR-Semi	0.11 (0.04)	1.00 (0.00)	1.00 (0.00)	0.20 (0.06)	0.99 (0.01)	0.99 (0.00)
IR-CP	0.14 (0.05)	0.99 (0.00)	1.00 (0.00)	0.27 (0.09)	0.98 (0.02)	0.99 (0.01)

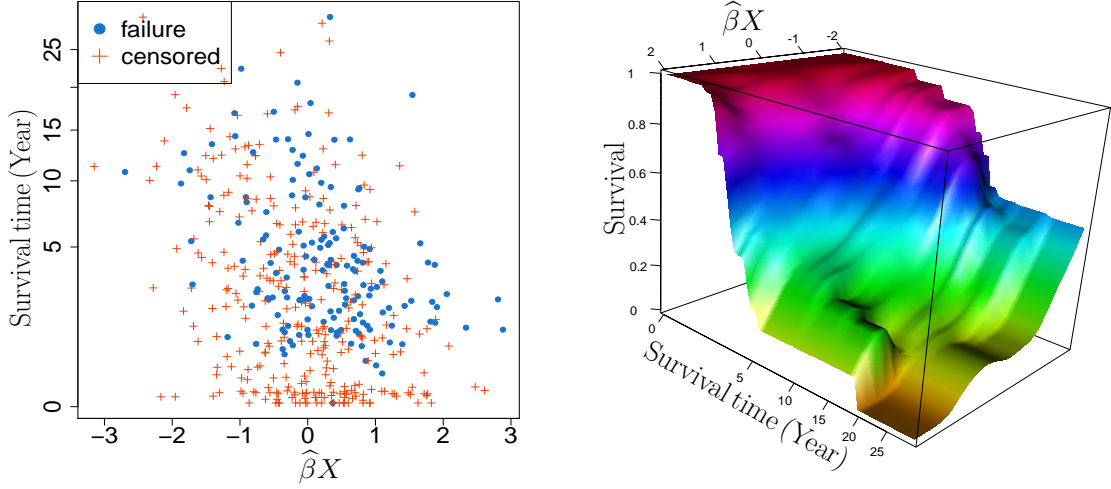
taneous melanoma based on a meta-analyses of over 145 existing literatures ([Chatzinasiou et al., 2011](#)). A list of these genes can be found at [www.melgene.org](http://www.melgene.org). We further include age at diagnosis as a clinical control variable.

Selecting the number of structural dimensions can be a challenging task, especially with right censored survival model ([Xia et al., 2010](#)). To this end, we adopt the validated information criterion (vic) developed by [Ma and Zhang \(2015\)](#), which is particularly suited for our GMM framework. The vic criterion is constructed by penalizing the quadratic form of the GMM approach (2.6). By minimizing  $\text{vic}(d)$ , the true structural dimension can be consistently chosen. Interestingly when we apply this method to all of our proposed estimating equation approaches,  $d = 1$  always yields the best fit. Hence we present the results for all method under  $d = 1$ . As a demonstration of the fitted model, we project the design matrix on the estimated direction of the Semi-CP approach and plot the survival outcome against the projection (Figure 1). A nonparametric estimation of the conditional survival function based on this projection is also produced. From these two figures, we can see a clear trend that subjects with larger values of the projection have lower survival rate.

As comparisons, we conduct all competing methods used in our simulation study with one structural dimension, and summarize all results in Figures 2 - 4 in Appendix B.

We observe both similarities and differences among different methods for the identified genes. A pairwise Frobenius norm distance is provided in Table 6 in Appendix B. This suggests that all proposed methods have fairly small distances, while existing methods do not agree with each other. In addition, DS seems to have the smallest distance with the proposed methods. The proposed methods share some consistent trend of the loadings on the most influential variables. For example, we identify Age as the most important variable with loading over 0.5. We then plot the observed data and also perform a nonparametric estimation of the conditional survival function based on the derived direction  $\hat{B}^T X$  for the IR-Semi approach (Figure 6). This suggests that higher survival rate is observed for smaller derived direction. This further indicates that patients with younger age tend to have higher survival, which is biologically intuitive. This finding is also consistent with the DS method, which identifies Age with loading 0.47. However, hMave and SIR-ipcw do not assign large loading to Age. Here, we present all results with positive loading of age, and the directions are multiplied by  $-1$  otherwise. Another important variable that all variables agrees (with both signs and magnitude) is CASP8. We mine the literature and found that Li et al. (2008) genotyped putatively functional polymorphisms of CASP8 and found significant association with lower risk of cutaneous melanoma. The result therein supports the large negative loading of CASP8 gene in our fitted model. Another gene that all method agrees with signs is CDKN2A. CDKN2A is a well documented gene for its association with hereditary melanoma. Individuals with a mutation in this gene (hence not properly expressed) may have up to a 67% lifetime risk for melanoma (Bishop et al., 2002). This coincide with the negative sign of all methods, although the magnitude is not among the largest ones. For differences across methods, hMave identifies TYR with largest loading (-0.79), which is not identified by other method as important factor. The enzyme

Figure 1: Survival Trend with Estimated Dimension Reduction Space.



The left figure is the projected direction versus the observed failure and censoring times. The right figure is a nonparametric estimation of the survival function based on the projected direction. In either case, a clear trend of increased hazard is observed for larger projection values.

encoded by this gene controls the production of melanin, hence it has been shown to be strongly associated with melanoma ([Gudbjartsson et al., 2008](#)). However, we were not able to confirm the sign of this gene. After conducting a plot and a nonparametric estimation of the survival function based on the hMave direction, it seems hard to judge whether the direction should be reverted.

## 6 Discussion

In this paper, we proposed a counting process based dimension reduction framework for censored outcomes. A family of generalized method of moments based approaches have been constructed for estimating the dimension reduction subspace. The main advantage of the proposed method is that it requires only a  $d$ -dimensional (instead of  $p$ ) nonparametric kernel estimation while no censoring distribution is modeled. The reduced dimension of the nonparametric estimation circumvents the difficulties of many exiting methods and improves

Table 5: SKCM data analysis results: the loading vectors of first structural dimension.

Covariate	DS	hMave	SIR-ipcw	FR	IR-Semi	IR-CP	CP-SIR
Age	0.47	0.00	0.06	0.55	0.54	0.59	0.59
TYRP1	-0.05	0.24	0.00	0.19	0.16	0.04	0.11
OCA2	0.17	-0.06	-0.10	0.18	0.19	0.23	0.19
TYR	-0.09	-0.79	0.30	-0.17	-0.08	-0.10	-0.27
SLC45A2	0.24	0.14	-0.04	0.31	0.31	0.32	0.28
CDKN2A	-0.28	-0.12	-0.18	-0.10	-0.11	-0.10	-0.07
MX2	-0.02	-0.12	0.00	-0.20	-0.19	-0.20	-0.14
MTAP	-0.08	-0.14	0.07	-0.30	-0.31	-0.34	-0.35
MITF	-0.09	0.05	-0.33	-0.21	-0.33	-0.25	-0.12
VDR	-0.18	0.10	0.00	-0.11	-0.04	-0.04	-0.06
CCND1	0.35	-0.05	0.43	0.21	0.13	0.13	0.17
MYH7B	-0.27	-0.04	-0.17	-0.32	-0.34	-0.33	-0.32
ATM	-0.22	0.28	-0.14	-0.02	0.07	0.06	0.07
PLA2G6	-0.16	0.07	0.00	0.06	0.10	0.05	-0.05
CASP8	-0.39	-0.13	-0.19	-0.24	-0.23	-0.27	-0.24
AFG3L1	0.26	-0.15	-0.07	0.12	0.02	0.13	0.10
CDK10	0.08	0.25	0.01	-0.06	-0.03	-0.05	-0.01
PARP1	0.03	0.17	0.45	0.18	0.19	0.14	0.18
CLPTM1L	-0.05	0.02	0.20	-0.05	-0.11	-0.10	-0.06
ERCC5	0.25	-0.07	0.43	0.14	0.12	0.03	0.13
FTO	-0.03	0.14	0.20	0.16	0.13	0.10	0.17

the efficiency when the total dimension  $p$  too large for kernel method. Our simulation study suggests that the proposed method outperforms existing methods in a variety of settings. The disadvantage of the proposed method is the additional computational cost. Solving the estimating equations with a large number of  $p$  can still be computationally intense. On the other hand, while our computational efficient approach requires only a singular value decomposition, and has satisfactory performances, it does not enjoy the same theoretical guardant without restraint conditions on the covariates. Further relaxation of these conditions is of great interest.

Our framework can be possibly extended to more general settings. First, by imposing penalization on the estimating equations, it is possible to extend the proposed method to moderately high dimensional data. Sparse estimation of the  $B$  parameter may help both interpretation and improves prediction accuracy of subsequent nonparametric models. The second direction is to search for alternative construction of the  $\alpha$  functions. Throughout our developments, we used the  $\phi(u)$  function which is motivated from the inverse regression of a binomial distribution. It would be interesting to investigate the possibilities of a “SAVE” type of  $\alpha$  function that may detect more complicated model structure. Lastly, it is also interesting to extend this framework to time varying coefficient setting, where we may let the dimension reduction space  $\mathcal{S}$  to change over time  $t$ .



## A Appendix

We collect technical details of Sections 2 and 4 in this appendix. We start with the derivation of the local nuisance tangent space and the global tangent space (2.1) in the following subsection.

### A.1 Tangent Space Derivation

Before we give the tangent space, we first derive the nuisance tangent spaces  $\mathcal{E}_1, \mathcal{E}_2$  and  $\mathcal{E}_3$  in the following proposition. The proof follows the similar argument for proving the nuisance tangent spaces of cox regression model (Tsiatis, 2007) and is thus omitted for simplicity.

**Proposition A.1.** The nuisance tangent spaces  $\mathcal{E}_1, \mathcal{E}_2$  and  $\mathcal{E}_3$  have the following forms

$$\begin{aligned}\mathcal{E}_1 &= \left\{ \int \alpha(u, B^T X) dM(u, B^T X) \text{ for all functions } \alpha(u, B^T X) \right\}, \\ \mathcal{E}_2 &= \left\{ \int \alpha(u, X) dM_C(u, X) \text{ for all functions } \alpha(u, X) \right\}, \text{ and} \\ \mathcal{E}_3 &= \left\{ \alpha(X) : E\{\alpha(X)\} = 0 \right\}.\end{aligned}$$

Next, we give a proof of tangent space, (2.1).

*Proof of (2.1).* For a fully nonparametric model, the nuisance tangent space is the whole Hilbert space  $\mathcal{H}$  with each element having mean zero. Therefore, if we put no restriction on the hazard function  $\lambda(t|X)$  and write the associated nuisance tangent space as  $\mathcal{E}_1^*$ , we obtain

$$\begin{aligned}\mathcal{H} &= \mathcal{E}_1^* \oplus \mathcal{E}_2 \oplus \mathcal{E}_3, \text{ where} \\ \mathcal{E}_1^* &= \left\{ \int \alpha(u, X) dM(u, X) : \text{for all functions } \alpha(u, X) \right\}.\end{aligned}$$

The orthogonal completion of  $\mathcal{E}$  satisfies that  $\mathcal{E}^\perp \subset \mathcal{E}_1^*$  and  $\mathcal{E}^\perp \perp \mathcal{E}_1$ . In order to identify  $\mathcal{E}^\perp$ , it suffices to take an arbitrary element in  $\mathcal{E}_1^*$  and find its residual after projecting it

onto  $\mathcal{E}_1$ . To find the projection, we must derive  $\alpha^*(u, B^T X) \in \mathcal{E}_1$  such that

$$E\left(\int (\alpha(u, X) - \alpha^*(u, B^T X))^T dM(u, X) \int a(u, B^T X) dM(u, X)\right) = 0.$$

The covariance of martingale stochastic integrals above can be computed by finding the expectation of the predictable covariance process ([Fleming and Harrington, 2011](#)):

$$\begin{aligned} & E\left(\int \{\alpha(u, X) - \alpha^*(u, B^T X)\}^T dM(u, X) \int a(u, B^T X) dM(u, X)\right) \\ &= E\left(\int \left[E\{\alpha(u, X) | \mathcal{F}_u, B^T X\} - \alpha^*(u, B^T X)\right]^T a(u, B^T X) \lambda(u | B^T X) Y(u) du\right) = 0, \end{aligned}$$

where  $a(u, B^T X) \in \mathcal{E}_1$  is arbitrary. Thus we must have

$$\alpha^*(u, B^T X) = E\{\alpha(u, X) | \mathcal{F}_u, B^T X\}.$$

This completes the proof. □

## A.2 Double Robustness Property for IR-Semi

In this section, we show that the estimator in IR-Semi enjoys the double robustness property.

Recall that, for IR-Semi, we solve the sample version of the following estimating functions

$$E\left[\int \left\{E(X|Y(u)) - E(X|Y(u), B^T X)\right\} \varphi^T(u) dM(u)\right] = 0.$$

For simplicity, we will use the random function to denote

$$F(X, u) = E(X|Y(u)) - E(X|Y(u), B^T X) \varphi^T(u).$$

**Case 1:** Suppose  $M(u)$  is misspecified as  $M^*(u)$ . Then we have that

$$\begin{aligned} E\{F(X, u) dM^*(u)\} &= E[E\{F(X, u) dM^*(u) | Y(u), X\}] = E[F(X, u) E\{dM^*(u) | Y(u), X\}] \\ &= E[E\{F(X, u) | Y(u), B^T X\} E\{dM^*(u) | Y(u), B^T X\}] = 0, \end{aligned}$$

where the last equation is due to the fact that  $E\{F(X, u) | Y(u), B^T X\} = 0$ . Hence we have

$$E\left[\int F(X, u) dM^*(u)\right] = 0.$$

**Case 2:** Suppose the function  $F(X, u)$  is misspecified to  $F^*(X, u)$ . In a similar argument, we can show that  $E\left\{\int F^*(X, u) dM(u)\right\} = 0$ . This completes the proof.

### A.3 Asymptotic Theory

Recall that  $\text{vec}(\cdot)$  is the concatenating operator and  $\beta = \text{vec}(B)$ . Let  $\Omega(B) = \{\tilde{B} : \|\tilde{B} - \beta\|_2 \leq Cn^{-1/2}\}$  and  $\Omega(\beta) = \{\tilde{\beta} : \|\tilde{\beta} - \beta\|_2 \leq Cn^{-1/2}\}$  for some  $C > 0$ . We need the following lemma, which is due to Lengart ([Fleming and Harrington, 2011](#)).

**Lemma A.1** (Lengart's Inequality). Suppose  $H$  is a predictable and locally bounded process. Then for any stopping time  $\tau$  such that  $\text{pr}(\tau < \infty) = 1$ , and any  $\epsilon, \eta > 0$ ,

$$\text{pr}\left[\sup_{t \leq \tau} \left\{ \int_0^t H(u) dM(u) \right\}^2 \geq \epsilon\right] \leq \frac{\eta}{\epsilon} + \text{pr}\left\{ \int_0^\tau H^2(u) d\langle M, M \rangle(u) \geq \eta \right\}.$$

Let  $G(u, B, X) = \text{vec}\{\alpha(u, X) - \alpha^*(u, B^\top X)\} \{\text{vec}(X \nabla_2 \lambda(u, B^\top X))\}^\top$  and  $A(u, B, X) = \text{vec}\{\alpha(u, X) - \alpha^*(u, B^\top X)\} [\text{vec}\{\alpha(u, X) - \alpha^*(u, B^\top X)\}]^\top$ . Then we are ready to prove the main theorem of this paper.

*Proof of Theorem 4.6.* Without loss of generality, we prove the theorem under the generic case in which  $\alpha^*(u, B^\top X)$  is arbitrary. For simplicity, we sometimes write  $M(u, B^\top X_i)$  as  $M_i(u)$ . Let

$$\begin{aligned} \widehat{S}_n(\widehat{B}) &= \frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \widehat{\alpha}^*(u, \widehat{B}^\top X_i) \right\} d\widehat{M}(u, \widehat{B}^\top X_i) \right] \\ &= \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \widehat{\alpha}^*(u, \widehat{B}^\top X_i) \right\} d\left\{ \widehat{M}(u, \widehat{B}^\top X_i) - M(u, B^\top X_i) \right\} \right]}_{\text{I}} \\ &\quad + \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \widehat{\alpha}^*(u, \widehat{B}^\top X_i) \right\} dM(u, B^\top X_i) \right]}_{\text{II}} \\ &\quad + \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \widehat{\alpha}^*(u, \widehat{B}^\top X_i) \right\} d\left\{ M(u, \widehat{B}^\top X_i) - M(u, B^\top X_i) \right\} \right]}_{\text{III}}. \end{aligned} \tag{A.1}$$

For simplicity, we sometimes write  $S_n(B)$  and  $S_n^*(B)$  as  $S_n(\beta)$  and  $S_n^*(\beta)$ , respectively.

Following [Jureckova \(1971\)](#) and [Tsiatis \(1990\)](#), it suffices to show that  $S_n(\beta)$  is uniformly

asymptotically linear in a small neighborhood of the true parameter  $\beta$ . That is, we need to show, there exists some linear operator,  $G_n$ , such that

$$\sup_{\|\hat{\beta}-\beta\|_2 \leq Cn^{-1/2}} n^{1/2} \|S_n(\hat{\beta}) - S_n(\beta) - G_n(\hat{\beta} - \beta)\|_2 = o_p(1).$$

To achieve such, we bound terms I, II and III, respectively. We start by decomposing I as follows.

$$\begin{aligned} \text{I} = & \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} d \left\{ \widehat{M}(u, \widehat{B}^T X_i) - M(u, \widehat{B}^T X_i) \right\} \right]}_{\text{I}_1} \\ & + \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha^*(u, B^T X_i) - \widehat{\alpha}^*(u, \widehat{B}^T X_i) \right\} d \left\{ \widehat{M}(u, \widehat{B}^T X_i) - M(u, \widehat{B}^T X_i) \right\} \right]}_{\text{I}_2} \\ & + \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} d \left\{ M(u, \widehat{B}^T X_i) - M(u, B^T X_i) \right\} \right]}_{\text{I}_3} \\ & + \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha^*(u, B^T X_i) - \widehat{\alpha}^*(u, \widehat{B}^T X_i) \right\} d \left\{ M(u, \widehat{B}^T X_i) - M(u, B^T X_i) \right\} \right]}_{\text{I}_4} \end{aligned}$$

For term  $\text{I}_1$ , using integration by parts along with Assumption 4.4 obtains us that

$$\begin{aligned} & \frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} Y_i(u) d \left\{ \widehat{\Lambda}(u, \widehat{B}^T X_i) - \Lambda(u, \widehat{B}^T X_i) \right\} \right] \\ &= \frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \left\{ \alpha(\tau, X_i) - \alpha^*(\tau, B^T X_i) \right\} Y_i(\tau) \left\{ \widehat{\Lambda}(\tau, \widehat{B}^T X_i) - \Lambda(\tau, \widehat{B}^T X_i) \right\} \right] \\ & \quad - \frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \widehat{\Lambda}(u, \widehat{B}^T X_i) - \Lambda(u, \widehat{B}^T X_i) \right\} Y_i(u) d \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} \right] \\ &= \frac{1}{n} \sum_{i=1}^n A_1(\tau, X_i) + \frac{1}{n} \left\{ \sum_{i=1}^n A_2(\tau, X_i) \right\} (\widehat{\beta} - \beta) + O_p(\|\widehat{\beta} - \beta\|_2^2) = O_p(n^{-1/2}), \end{aligned}$$

where  $A_1(\tau, X_i) = \text{vec} \left\{ \alpha(\tau, X_i) - \alpha^*(\tau, B^T X_i) \right\} Y_i(\tau) \psi_1(t, B^T X_i) - \int_0^\tau \psi_1(u, B^T X_i) Y_i(u) d \text{vec} \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\}$ , and  $A_2(\tau, X_i) = \text{vec} \left\{ \alpha(\tau, X_i) - \alpha^*(\tau, B^T X_i) \right\} Y_i(\tau) \psi_2(t, B^T X_i)^T - \int_0^\tau Y_i(u) d \text{vec} \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} \psi_2(u, B^T X_i)$ . For the second term, using Assumption 4.3 and Assumption 4.4,

we have

$$\begin{aligned}
I_2 &= \frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha^*(u, B^T X_i) - \alpha^*(u, \widehat{B}^T X_i) \right\} d \left\{ \widehat{M}(u, \widehat{B}^T X_i) - M(u, \widehat{B}^T X_i) \right\} \right] \\
&\quad + \frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha^*(u, \widehat{B}^T X_i) - \widehat{\alpha}^*(u, \widehat{B}^T X_i) \right\} d \left\{ \widehat{M}(u, \widehat{B}^T X_i) - M(u, \widehat{B}^T X_i) \right\} \right] \\
&= O_p(n^{-3/2+\kappa/2} + n^{-1+\kappa}) = o_p(n^{-1/2}).
\end{aligned}$$

In a similar argument, we can show that  $I_4 = O_p(n^{-1} + n^{-3/2+\kappa/2}) = o_p(n^{-1/2})$ . We use  $\nabla_2 \lambda(u, a)$  to denote the  $k$ -dimensional sub-gradient vector of  $\lambda(u, a)$  with respect to the second argument at  $(u, a)$ . Then

$$\begin{aligned}
I_3 &= \frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} \text{tr} \left[ X_i \left\{ \nabla_2 \lambda(u, B^T X_i) \right\}^T (\widehat{B} - B) \right] du \right] + O_p(\|\widehat{\beta} - \beta\|_2^2) \\
&= \frac{1}{n} \left[ \sum_{i=1}^n \int_0^\tau \text{vec} \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} \left\{ \text{vec} \left( X_i \nabla_2 \lambda(u, B^T X_i) \right) \right\}^T du \right] (\widehat{\beta} - \beta) + O_p\left(\frac{1}{n}\right).
\end{aligned}$$

Next, we give the asymptotically linear expansion for term II:

$$\begin{aligned}
II &= \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} dM_i(u) \right]}_{\Pi_1} \\
&\quad + \underbrace{\frac{1}{n} \text{vec} \left[ \sum_{i=1}^n \int_0^\tau \left\{ \alpha^*(u, \widehat{B}^T X_i) - \widehat{\alpha}^*(u, \widehat{B}^T X_i) \right\} dM_i(u) \right]}_{\Pi_2}.
\end{aligned}$$

Using Assumptions 4.1 and 4.4, we obtain that, there exists a  $\kappa < 1/2$  such that

$$\sup_{0 \leq u \leq \tau} \left| \text{vec} \left\{ \widehat{\alpha}^*(u, \widehat{B}^T X) - \alpha^*(u, \widehat{B}^T X) \right\} \right| = o_p(n^{-1/2+\kappa/2}).$$

This further yields that

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n \text{vec} \left[ \int_0^\tau \left\{ \widehat{\alpha}^*(u, \widehat{B}^T X_i) - \alpha^*(u, \widehat{B}^T X_i) \right\}^2 Y_i(u) d\Lambda_i(u) \right] \\
&\leq \frac{1}{n} \sum_{i=1}^n \text{vec} \left[ \Lambda_i(\tau) \sup_{0 \leq u \leq \tau} \left\{ \widehat{\alpha}^*(u, \widehat{B}^T X_i) - \alpha^*(u, \widehat{B}^T X_i) \right\}^2 \right] \leq o_p(n^{-1+\kappa}) = o_p(n^{-1/2}).
\end{aligned}$$

Therefore, Lengart's inequality in Lemma A.1 implies that

$$\begin{aligned} & \Pr \left\{ \sup_{0 \leq t \leq \tau} \left( \text{vec} \left[ n^{-1/2} \sum_{i=1}^n \int_0^t \{ \hat{\alpha}^*(u, \hat{B}^T X) - \alpha^*(u, \hat{B}^T X) \} dM_i(u) \right] \right)^2 \leq \epsilon \right\} \\ & \leq \frac{\eta}{\epsilon} + \Pr \left( \text{vec} \left( n^{-1/2} \sum_{i=1}^n \int_0^t \{ \hat{\alpha}^*(u, \hat{B}^T X_i) - \alpha^*(u, \hat{B}^T X_i) \} \right)^2 Y_i(u) d\Lambda_i(u) \right) > \eta \right). \end{aligned}$$

Taking  $\eta \rightarrow 0$ , we obtain that  $\sqrt{n} \text{II}_2 = o_p(1)$ . Last, we bound term III. In a similar argument for bounding I, we obtain

$$\text{III} = \frac{1}{n} \left[ \sum_{i=1}^n \int_0^\tau \text{vec} \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} \left\{ \text{vec} \left( X_i \nabla_2 \lambda(u, B^T X_i) \right) \right\}^T du \right] (\hat{\beta} - \beta) + O_p(n^{-1+\kappa/2}).$$

Combing the asymptotically linear expansion for terms I, II and III.  $\hat{S}_n(\hat{B})$  can be written as

$$\begin{aligned} \hat{S}_n(\hat{B}) &= \frac{1}{n} \sum_{i=1}^n \left[ A_1(\tau, X_i) + \int_0^\tau \text{vec} \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} dM_i(u) \right] \\ &\quad \underbrace{\hspace{10em}}_{S_n(B)} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left\{ A_2(\tau, X_i) + 2 \int_0^\tau G(u, B, X_i) du \right\} (\beta - \hat{\beta}) + o_p(n^{-1/2}) = 0, \\ &\quad \underbrace{\hspace{10em}}_{G_n} \end{aligned}$$

in which  $G(u, B, X) = \text{vec} \left\{ \alpha(u, X) - \alpha^*(u, B^T X) \right\} \left\{ \text{vec} \left( X \nabla_2 \lambda(u, B^T X) \right) \right\}^T$ . Therefore,

$\hat{\beta} - \beta$  can be written as

$$\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n} G_n^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left[ A_2(\tau, X_i) + \int_0^\tau \text{vec} \left\{ \alpha(u, X_i) - \alpha^*(u, B^T X_i) \right\} dM_i(u) \right] \right).$$

This finishes the proof. □

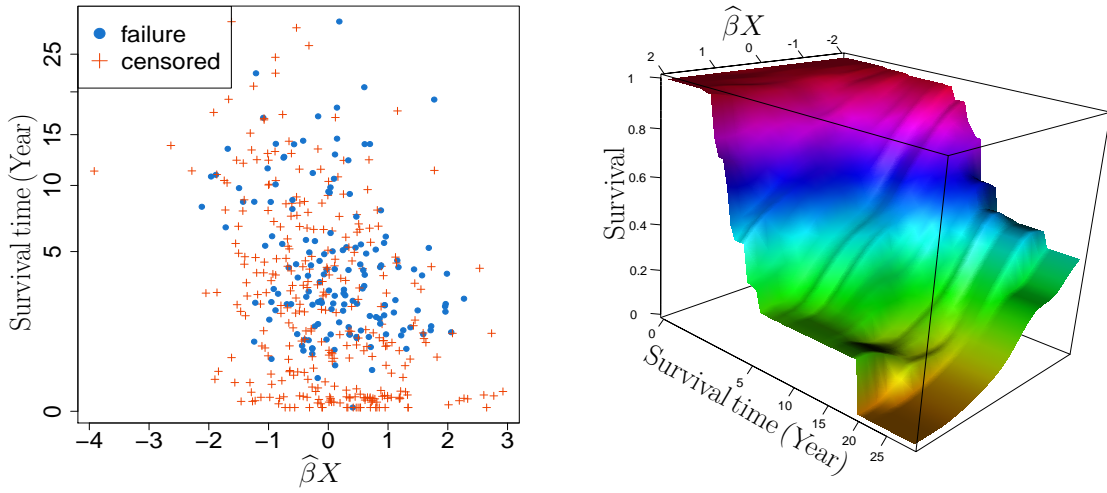
## B Additional Results

This section contains additional results in the SKCM data analysis. Table 6 is the pairwise distance measure of the first direction estimated by all methods in the SKCM data. This shows that the proposed methods mostly agree with each other. Figures 2 - 4 are the fitted models of competing approaches.

Table 6: Pairwise distance measure among all methods

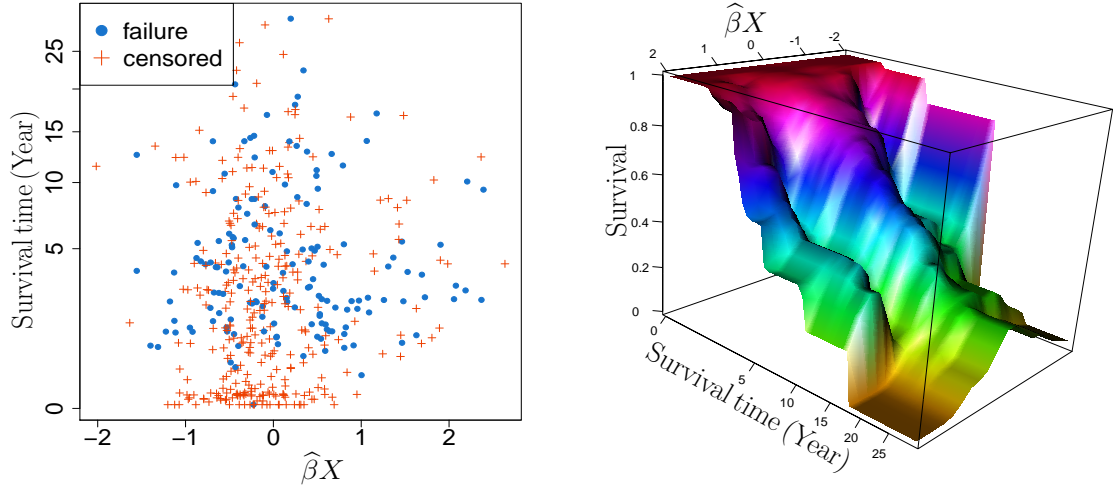
	DS	hMave	SIR-ipcw	FR	IR-Semi	IR-CP
hMave	1.41					
SIR-ipcw	1.27	1.39				
FR	0.89	1.35	1.31			
IR-Semi	0.94	1.37	1.36	0.38		
IR-CP	1.02	1.36	1.31	0.35	0.34	
CP-SIR	0.92	1.29	1.36	0.35	0.42	0.50

Figure 2: Survival Trend with Estimated Dimension Reduction Space (DS).



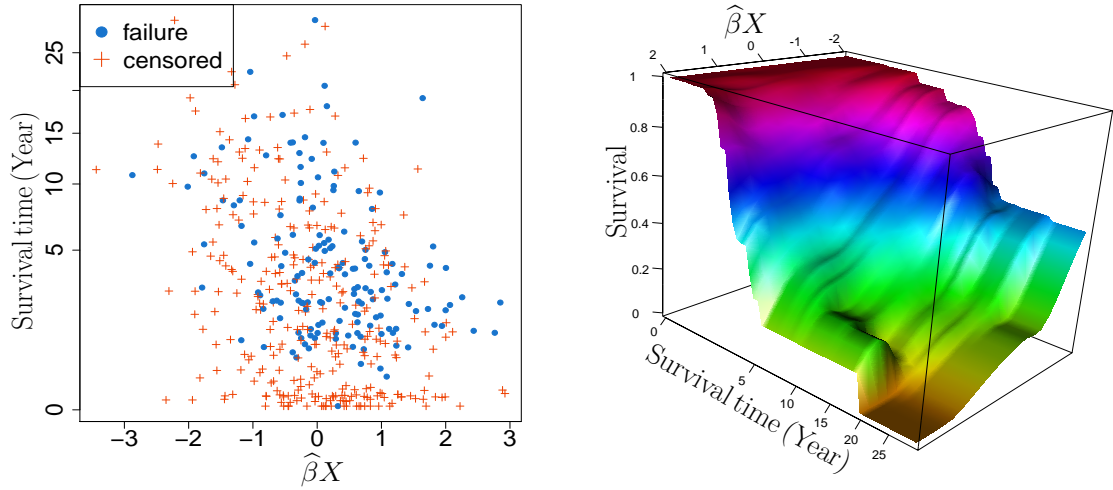
The left figure is the projected direction versus the observed failure and censoring times. The right figure is a nonparametric estimation of the survival function based on the projected direction. In either case, a clear trend of increased hazard is observed for larger projection values.

Figure 3: Survival Trend with Estimated Dimension Reduction Space (hMave).



The left figure is the projected direction versus the observed failure and censoring times. The right figure is a nonparametric estimation of the survival function based on the projected direction. In either case, a clear trend of increased hazard is observed for larger projection values.

Figure 4: Survival Trend with Estimated Dimension Reduction Space (SIR-ipcw).



The left figure is the projected direction versus the observed failure and censoring times. The right figure is a nonparametric estimation of the survival function based on the projected direction. In either case, a clear trend of increased hazard is observed for larger projection values.



## References

- Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press Baltimore.
- Bishop, D. T., Demenais, F., Goldstein, A. M., Bergman, W., Bishop, J. N., Bressac-de Paillerets, B., Chompret, A., Ghiorzo, P., Gruis, N., Hansson, J., et al. (2002), “Geographical variation in the penetrance of CDKN2A mutations for melanoma,” *Journal of the National Cancer Institute*, 94, 894–903.
- Chatzinasiou, F., Lill, C. M., Kypreou, K., Stefanaki, I., Nicolaou, V., Spyrou, G., Evangelou, E., Roehr, J. T., Kodela, E., Katsambas, A., et al. (2011), “Comprehensive field synopsis and systematic meta-analyses of genetic association studies in cutaneous melanoma,” *Journal of the National Cancer Institute*, 103, 1227–1235.
- Cook, R. D. (2009), *Regression Graphics: Ideas for Studying Regressions Through Graphics*, John Wiley & Sons.
- Cook, R. D. and Lee, H. (1999), “Dimension reduction in binary response regression,” *Journal of the American Statistical Association*, 94, 1187–1200.
- Cook, R. D. and Weisberg, S. (1991), “Discussion of sliced inverse regression for dimension reduction,” *Journal of the American Statistical Association*, 86, 328–332.
- Cox, D. R. (1972), “Regression models and life-tables,” *Journal of the Royal Statistical Society: Series B*, 187–220.
- Dabrowska, D. M. et al. (1989), “Uniform consistency of the kernel conditional Kaplan-Meier estimate,” *The Annals of Statistics*, 17, 1157–1167.
- Delyon, B., Portier, F., et al. (2016), “Integral approximation by kernel smoothing,” *Bernoulli*, 22, 2177–2208.

- Dong, Y. and Li, B. (2010), “Dimension reduction for non-elliptically distributed predictors: second-order methods,” *Biometrika*, 97, 279–294.
- Fleming, T. R. and Harrington, D. P. (2011), *Counting Processes and Survival Analysis*, John Wiley & Sons.
- Gudbjartsson, D. F., Sulem, P., Stacey, S. N., Goldstein, A. M., Rafnar, T., Sigurgeirsson, B., Benediktsdottir, K. R., Thorisdottir, K., Ragnarsson, R., Sveinsdottir, S. G., et al. (2008), “ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma,” *Nature genetics*, 40, 886–891.
- Hansen, L. P. (1982), “Large sample properties of generalized method of moments estimators,” *Econometrica*, 1029–1054.
- Jureckova, J. (1971), “Nonparametric estimate of regression coefficients,” *The Annals of Mathematical Statistics*, 1328–1338.
- Li, B. and Dong, Y. (2009), “Dimension reduction for nonelliptically distributed predictors,” *The Annals of Statistics*, 1272–1298.
- Li, B. and Wang, S. (2007), “On directional regression for dimension reduction,” *Journal of the American Statistical Association*, 102, 997–1008.
- Li, C., Zhao, H., Hu, Z., Liu, Z., Wang, L.-E., Gershenwald, J. E., Prieto, V. G., Lee, J. E., Duvic, M., Grimm, E. A., et al. (2008), “Genetic variants and haplotypes of the caspase-8 and caspase-10 genes contribute to susceptibility to cutaneous melanoma,” *Human Mutation*, 29, 1443–1451.
- Li, K.-C. (1991), “Sliced inverse regression for dimension reduction,” *Journal of the American Statistical Association*, 86, 316–327.

- Li, K.-C., Wang, J.-L., and Chen, C.-H. (1999), “Dimension reduction for censored regression data,” *The Annals of Statistics*, 27, 1–23.
- Lin, D., Wei, L., and Ying, Z. (1998), “Accelerated failure time models for counting processes,” *Biometrika*, 85, 605–618.
- Lu, W. and Li, L. (2011), “Sufficient dimension reduction for censored regressions,” *Biometrics*, 67, 513–523.
- Ma, Y. and Zhang, X. (2015), “A validated information criterion to determine the structural dimension in dimension reduction models,” *Biometrika*, 102, 409–420.
- Ma, Y. and Zhu, L. (2012), “A semiparametric approach to dimension reduction,” *Journal of the American Statistical Association*, 107, 168–179.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, CRC press.
- Sun, W. and Wu, T. (2006), *censorSIR*, tools for applying Slice Inverse Regression in censored data.
- Tsiatis, A. (2007), *Semiparametric Theory and Missing Data*, Springer Science & Business Media.
- Tsiatis, A. A. (1990), “Estimating regression parameters using linear rank tests for censored data,” *The Annals of Statistics*, 354–372.
- Weisberg, S. (2002), “Dimension reduction regression in R,” *Journal of Statistical Software*, 7, 1–22.
- Xia, Y. (2007), “A constructive approach to the estimation of dimension reduction directions,” *The Annals of Statistics*, 35, 2654–2690.
- Xia, Y., Zhang, D., and Xu, J. (2010), “Dimension reduction and semiparametric estimation of survival models,” *Journal of the American Statistical Association*, 105, 278–290.

Zeng, D. and Lin, D. (2007), “Maximum likelihood estimation in semiparametric regression models with censored data,” *Journal of the Royal Statistical Society: Series B*, 69, 507–564.

Zhu, L., Miao, B., and Peng, H. (2006), “On sliced inverse regression with high-dimensional covariates,” *Journal of the American Statistical Association*, 101, 630–643.